Amendments to the Specification

Please replace the section "**BRIEF DESCRIPTION OF THE DRAWINGS**" from page 10, line 11, to page 28, line 2, with the following replacement section.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate several embodiments of the disclosed method and compositions and together with the description, serve to explain the principles of the disclosed method and compositions.

Figures 1A and 1B show metabolite-dependent conformational changes in the 202-nucleotide leader sequence of the *btuB* mRNA. Figure 1A shows separation of spontaneous RNA-cleavage products of the *btuB* leader using denaturing 10% polyacrylamide gel electrophoresis (PAGE). 5'-32p-labeled mRNA leader molecules (arrow) were incubated for 41 hr at 25°C in 20 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.3 at 25°C) in the presence (+) or absence (-) of 20 μM of AdoCbl. Lanes containing RNAs that have undergone no reaction, partial digest with alkali, and partial digest with RNase T1 (G-specific cleavage) are identified by NR, ^{-}OH, and T1, respectively. The location of product bands corresponding to cleavage after selected guanosine residues are identified by filled arrowheads. ~~Light blue arrowheads~~ Arrowheads labeled 1 through 8 identify eight of the nine locations that exhibit effector-induced structure modulation, which experience an increase or decrease in the rate of spontaneous RNA cleavage. The image was generated using a phosphorimager (Molecular Dynamics), and cleavage yields were quantitated by using ImageQuant software. Figure 1B shows sequence and secondary-structure model for the 202-nucleotide leader sequence of *btuB* mRNA (SEQ ID NO: 1) in the presence of AdoCbl. Putative base-paired elements are designated P1 through P9. Complementary nucleotides in the loops of P4 and P9 that have the potential to form a pseudoknot are juxtaposed. Nine specific sites of structure modulation are identified by ~~light blue~~

arrowheads. The asterisks demark the boundaries of the B_{12} box (nucleotides 141-162). The coding region and the 38 nucleotides that reside immediately 5' of the start codon (nucleotides 241-243) were not included in the 202-nucleotide fragment. The 315-nucleotide fragment includes the 202-nucleotide fragment, the remaining 38 nucleotides of the leader sequence, and the first 75 nucleotides of the coding region.

Figures 2A and 2B show the *btuB* mRNA leader forms a saturable binding site for AdoCbl. Figure 2A shows the dependence of spontaneous cleavage of *btuB* mRNA leader on the concentration of AdoCbl effector as represented by site 1 (G23) and site 2 (U68). 5'-^{32}P-labeled mRNA leader molecules were incubated, separated, and analyzed as described in the in the brief description of Figure 1, and include identical control and marker lanes as indicated. Incubations contained concentrations of AdoCbl ranging from 10 nM to 100 μM (lanes 1 though 8) or did not include AdoCbl (-). Figure 2B shows a composite plot of the fraction of RNA cleaved at six locations along the mRNA leader versus the logarithm of the concentration (c) of AdoCbl. Fraction cleaved values were normalized relative to the highest and lowest cleavage values measured for each location, including the values obtained upon incubation in the absence of AdoCbl. The inset defines the symbols used for each of six sites, while the remaining three sites were excluded from the analysis due to weak or obscured cleavage bands. Filled and open symbols represent increasing and decreasing cleavage yields, respectively, upon increasing the concentration of AdoCbl. The dashed line reflects a K_D of ~300 nM, as predicted by the concentration needed to generate half-maximal structural modulation. Data plotted were derived from a single PAGE analysis, of which two representative sections are depicted in Figure 1A.

Figure 3 shows the 202-nucleotide mRNA leader causes an unequal distribution of AdoCbl in an equilibrium dialysis apparatus. I: Equilibration of tritiated effector was conducted in the absence of RNA. II: (step 1) Equilibration was conducted as in I, but with 200 pmoles of mRNA leader added to chamber b; (step 2) 5,000 pmoles of unlabeled AdoCbl was added to chamber b. III: Equilibrations were conducted as described in II, but wherein 5,000 pmoles of cyanocobalamin was added to chamber b. IV: (step 1) Equilibration was initiated as described in step 1 of II; (steps 2 and 3) the solution in chamber a was replaced with 25 μL of fresh equilibration buffer; (step 4) 5,000 pmoles of unlabeled AdoCbl was added to chamber *b*. The

cpm ratio is the ratio of counts detected in chamber b relative to that of *a*. The dashed line represents a cpm ratio of 1, which is expected if equal distribution of tritium is established.

Figures 4A and 4B show selective molecular recognition of effectors by the *btuB* mRNA leader. Figure 4A shows a chemical structure of AdoCbl (1) and various effector analogs (2 through 11, ref. 30). Figure 4B shows a determination of analog binding by monitoring modulation of spontaneous cleavage of the 202-nucleotide *btuB* RNA leader. 5'-^{32}P-labeled mRNA leader molecules were incubated, separated, and analyzed as described in the legend to Figure 1A, and include identical control and marker lanes as indicated. The sections of three PAGE analyses encompassing site 2 (U68) are depicted. Below each image is plotted the amount of RNA cleaved (normalized with relation to the lowest and highest levels of cleavage at U68 in each gel) for each effector as indicated, or for no effector (-). The compound 11 (13-epi-AdoCbl) is an epimer of AdoCbl wherein the configuration at C13 is inverted, so that the e propionamide side chain is above the plane of the corrin ring; see Brown et al., Conformational studies of 5′-deoxyadenosyl-13-epicobalamin, a coenzymatically active structural analog of coenzyme B_{12}. Polyhedron 17, 2213 (1998).

Figures 5A, 5B, 5C, 5D, 5E and 5F show mutations in the mRNA leader and their effects on AdoCb1 binding and genetic control. Figure 5A shows sequence of the putative P5 element of the wild-type 202-nucleotide *btuB* leader exhibits AdoCbl-dependent modulation of structure as indicated by the observed increase in spontaneous RNA cleavage at position U68 (10% denaturing PAGE gel). Assays were conducted in the absence (-) or presence (+) of 5 μM AdoCbl. The remaining lanes are as described in the legend to Figure 1A. The composite bar graph reflects the ability of the RNA to shift the equilibrium of AdoCbl in an equilibrium dialysis apparatus and the ability of a reporter gene (see Experimental Procedures) to be regulated by AdoCbl addition to a bacterial culture. (Left) Plotted is the cpm ratio derived by equilibrium dialysis, wherein chamber b contains the RNA. Details of the equilibrium dialysis experiments are described in the brief description of Figure 3. (Right) Plotted are the expression levels of β-galactosidase as determined from cells grown in the absence (-) or presence (+) of 5 μM AdoCbl. Boxed numbers on the left and right, respectively, reflect the approximate K_D and the fold repression of β-galactosidase activity in the presence of AdoCbl. N.D. designates not

determined. Figure 5B-5F shows sequences and performance characteristics of various mutant leader sequences as indicated. Constructs were created as described in the Experimental Procedures section.

Figures 6A, 6B, 6C and 6D show metabolite binding by mRNAs. Figure 6A shows TPP-dependent modulation of the spontaneous cleavage of 165 *thiM* RNA was visualized by polyacrylamide gel electrophoresis (PAGE). 5′ ^{32}P-labeled RNAs (arrow, 20 nM) were incubated for approximately 40 hr at 25°C in 20 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.3 at 25°C) in the presence (+) or absence (–) of 100 μM TPP. NR, $^{-}$OH and T1 represent RNAs subjected to no reaction, partial digestion with alkali, or partial digestion with RNase T1 (G-specific cleavage), respectively. Product bands representing cleavage after selected G residues are numbered and identified by filled arrowheads. The asterisk identifies modulation of RNA structure involving the Shine-Dalgarno (SD) sequence. Gel separations were analyzed using a phosphorimager (Molecular Dynamics) and quantitated using ImageQuant software. Figure 6B shows a secondary-structure model of 165 *thiM* (SEQ ID NO: 2) as predicted by computer modeling (Zuker et al., Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA Biochemistry and Biotechnology* (eds. Barciszewski J. & Clark, B.F.C.) 11-43 (NATO ASI Series, Kluwer Academic Publishers, 1999); Mathews et al., Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288, 911-940 (1999)) and by the structure probing data depicted in Figure 6A. Spontaneous cleavage characteristics are as noted in the inset. Unmarked nucleotides exhibit a constant but low level of degradation. The truncated 91 *thiM* RNA (residues 1-91 of SEQ ID NO: 2) is boxed and the *thi* box element (Miranda-Rios et al., A conserved RNA structure (*thi* box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc. Natl. Acad. Sci. USA* 98, 9736-9741 (2001)) is shaded ~~light blue~~. Nucleotides enclosed in boxes ~~highlighted in orange~~ identify an alternative pairing, designated P8*. The RNA carries two mutations (G156A and U157C) relative to wild type that were introduced in a non-essential portion of the construct to form a restriction site for cloning, while all RNAs carry two 5′-terminal G residues to facilitate *in vitro* transcription. Figure 6C shows TPP-dependent modulation of the spontaneous cleavage of 240 *thiC* RNA. Reactions were

conducted and analyzed as described in above for Figure 6A. Figure 6D shows a secondary-structure model of 240 *thiC* (SEQ ID NO: 3). Base-paired elements that are similar to those of *thiM* are labeled P1 through P5. The truncated RNA 111 *thiC* (residues 1-111 of SEQ ID NO: 3) is boxed. Nucleotides enclosed in boxes ~~highlighted in orange~~ identify an alternative pairing.

Figures 7A, 7B and 7C show the *thiM* and *thiC* mRNA leaders serve as high-affinity metabolite receptors. Figure 7A shows the extent of spontaneous modulation of RNA cleavage at several sites within 165 *thiM* (left) and 240 *thiC* (right) plotted for different concentrations (*c*) of TPP. ~~Red arrows~~ Arrows reflect the estimated concentration of TPP needed to attain half maximal modulation of RNA (apparent K_D). Figure 7B shows the logarithm of the apparent K_D values plotted for both RNAs with TPP, TP and thiamine as indicated. The boxed data was generated using TPP with the truncated RNAs 91 *thiM* and 111 *thiC*. Figure 7C shows that patterns of spontaneous cleavage of 165 *thiM* differ between thiamine and TPP ligands as depicted by PAGE analysis (left) and as reflected by graphs (right) representing the relative phosphorimager counts for the three lanes as indicated. Details for the RNA probing analysis are similar to those described above in connection with Figure 6A. The graphs were generated by ImageQuant software.

Figures 8A, 8B and 8C show high sensitivity and selectivity of mRNA leaders for metabolite binding. Figure 8A shows chemical structures of several analogues of thiamine. TD is thiamine disulfide and THZ is 4-methyl-5-β-hydroxyethylthiazole. Figure 8B shows PAGE analysis of 165 *thiM* RNA structure probing using TPP and various chemical analogues (40 μM each) as indicated. Locations of significant structural modulation within the RNA spanning nucleotides ~113 to ~150 are indicated by open arrowheads. The asterisk identifies the site (C144) used to compare the normalized fraction of RNA that is cleaved (bottom) in the presence of specific compounds. Details for the RNA probing analysis are similar to those described above in connection with Figure 6a. Figure 8C shows a summary of the features of TPP that are critical for molecular recognition. Figure 8D shows equilibrium dialysis using ^{3}H-thiamine as a tracer. Plotted are the ratios for tritium distribution in a two-chamber system (*a* and *b*) that were established upon equilibration in the presence of the RNA constructs in chamber *b* as indicated

(see below for a description of the non-TPP-binding mutant M3). 100 μM TPP or oxythiamine were added to chamber *a*, as denoted, upon the start of equilibration.

Figures 9A, 9B and 9C show mutational analysis of the structure and function of the *thiM* riboswitch. Figure 9A shows mutations present in constructs M1 through M8 relative to the 165 *thiM* RNA (SEQ ID NO: 4). P8* is a putative base-paired element between portions (~~orange~~ encircled) of the P1 and P8 stems. Figure 9B (top) shows *in vitro* ligand-binding and genetic control functions of the wild-type (WT), M1 and M2 RNAs as reflected by PAGE analysis of in-line probing experiments (10 μM TPP) and by β-galactosidase expression assays. Labels on PAGE gels are as described above in connection with Figure 6A. Bars represent the levels of gene expression in the presence (+) and the absence (-) of TPP in the culture medium. Figure 9C is a summary of similar analyses of WT through M9 is presented in table form. The SD status "n.d." (not determined) indicates either that the level of spontaneous cleavage detected in the absence and presence of TPP is near the limit of detection (M6, M7 and M8) or that the region adopts an atypical structure (M9) compared to WT.

Figure 10 shows a construct for the selection of SAM-responsive ribozymes (SEQ ID NO:5). The hammerhead self-cleaving ribozyme and the SAM aptamer both require proper formation of the bridge domain to exhibit function. Therefore, the selection is expected to permit ribozyme function only when SAM or another binding-competent analog is present.

Figures 11A (SEQ ID NO: 6 and SEQ ID NOS: 378-382), 11B (SEQ ID NO: 7 and SEQ ID NOS: 383-385), 11C (SEQ ID NO: 8 and SEQ ID NOS: 386-387), 11D (SEQ ID NO: 9 and SEQ ID NOS: 388-389), 11E (SEQ ID NO: 10), 11F (SEQ ID NO: 11) and 11G (SEQ ID NO: 12 and SEQ ID NOS: 390-397) show consensus sequences and putative secondary structures were derived by phylogenetic and biochemical analyses as described for each riboswitch (see references). Nucleotides ~~in red~~ identified by a lower case a, c, t, or g, are conserved in greater than 90% of the representative sequences, open circles identify nucleotide positions of variable sequence, and lines identify elements that are variable in sequence and length. Models are described as follows: 6A) coenzyme B12 aptamer (Example 1); 6B) TPP aptamer (Example 2); 6C) FMN aptamer (Example 3); 6D) SAM aptamer (Example 7); 6E) guanine aptamer (Example 6); 6F) adenine aptamer (Example 8); and 6G) lysine aptamer Example 5). Letters R and Y

represent purine and pyrimidine bases, respectively; K designates G or U; W designates A or U; H designates A, C, or U; D designates G, A, or U; N represents any of the four bases.

Figures 12A (SEQ ID NO: 13), 12B and 12C show the regulation of the B. subtilis ribD mRNA by FMN. Figure 12A shows the results of in-line probing assays. Internucleotide linkages identified with ~~red circles~~ squares exhibit decreased amounts of spontaneous cleavage when ribD is incubated in the presence of FMN (indicating an increase in order for these nucleotides) relative to incubation in the absence of FMN. ~~Yellow circles~~ Circles identify linkages that exhibit consistently high levels of scission, which indicates they are not modulated by presence of FMN. Figure 12B shows a model for the mechanism of ribD regulation. The ribD mRNA adopts anti-termination conformation in the absence of FMN. Increased levels of FMN stabilize an RFN-FMN complex that permits formation of the terminator structure. Figure 12C shows the chemical structure and apparent dissociation constants for riboflavin and FMN.

Figures 13A (residues 1-91 of SEQ ID NO: 2), 13B and 13C show the regulation of the E. coli thiM mRNA by TPP. Figure 13A shows results of in-line probing assays. Internucleotide linkages identified with ~~red circles~~ squares exhibit decreased amounts of spontaneous cleavage when thiM is incubated in the presence of TPP compared to incubation in the absence of ligand. In contrast, linkages identified with ~~green circles~~ hexagons exhibit increased amounts of cleavage when thiM is incubated with TPP compared to incubation in the absence of ligand. The boxed nucleotides indicate ~~Blue shaded box indicates~~ pyrophosphate-recognition region (as described in text). Figure 13B shows a model for the mechanism of thiM regulation. In the absence of TPP, the anti-SD sequence interacts with part of aptamer domain to form anti-anti-SD. As TPP is increased, aptamer-TPP complexes are formed and the anti-SD favors pairing with the SD. Figure 13C shows the chemical structure and apparent dissociation constants for thiamine and TPP.

Figures 14A, 14B and 14C show putative eukaryote riboswitches. Figure 14A shows the consensus TPP binding domain based on 100 bacteria and archaea RNAs (SEQ ID NO: 18 and SEQ ID NOS: 398-399). Nucleotides ~~in red~~ shown as lower case letters are most conserved (>90%). Open circles represent nucleotide positions and domains that vary in sequence and length are designated *var*. The consensus model is similar to that reported recently (Rodionov et

al., 2002). Figure 14B the TPP-binding domain of *A. thaliana*(SEQ ID NO: 14). Variations in *O. sativa* (~~orange~~ nucleotides enclosed in a circle) (SEQ ID NO: 15) and *P. secunda* (~~green~~ nucleotides enclosed in a hexagon) (SEQ ID NO: 16) are shown. Figure 14C shows a putative TPP-binding domain in the intron of *N. crassa* (SEQ ID NO: 17).

Figure 15 shows sequence alignments of eukaryotic domains related to bacterial TPP-dependent riboswitches, Eco1, Eco2, Cac1, Ncr1, Aor1, Fox1, Fso1, Ath1, Pse1, Osa1, which are represented by SEQ ID NO: 19-28 respectively. Base paired stems are shaded in black and labeled as defined in Example 2). The P3 sequences, which in eukaryotes are significantly expanded in length and number of base pairs, are represented as a stem-loop structure. The highly conserved nucleotide positions in bacteria that were used to search for eukaryotic domains are ~~shaded gray~~ enclosed in a box. For each identified (ID) sequence, the position of the conserved CUGAGA sequence within the given Genbank entry is given along with the accession identification, sequence name, and gene identification. Additional protein annotations based on sequence similarity are shown in brackets. Methods: Riboswitch-like domains were initially identified by sequence similarity to bacterial sequences (Eco2 and Cac) by a blastn search of Genbank using default parameters. These hits were verified and expanded by searching for degenerate matches to the pattern (CTGAGA [200] ACYTGA [5] <<< GNTNNNNC >>> [5] CGNRGGRA) (SEQ ID NO: 375). Angle brackets indicate base pairing and bracketed numbers are variable gaps with constrained maximum lengths. All of the eukaryotic sequences have one or zero mismatches to this pattern except for one (Aor) that initially had three mismatches due to a single A insertion in the final search element. This mutation was removed to simplify the alignment. Comparison of mRNA (M33643.1) and genomic (AB033416.1) sequences demonstrated that the *F. oxysporum* element is in an intron in the 5' UTR of the sti35 gene. Other fungal sequences (Ncr, Aor, and Fso) are flanked by consensus splicing sequences.

Figures 16A and 16B show the structural probing of the putative TPP-riboswitch from *Arabidopsis*. Figure 16A shows the fragmentation pattern of the 128-nucleotide RNA (arrow) of *A. thaliana* (Fig. 14B) which was generated by incubation in the absence (-) or presence (+) of 100 μM TPP. T1, $^{-}$OH and NR identify RNAs that were partially digested with RNase T1 (cleaves 3′ to G residues), alkali, or were not reacted, respectively. Reactions were conducted as

described in Example 2. Figure 16B shows the apparent K_D for TPP binding by the *A. thaliana* RNA. Fraction bound was determined by in-line probing as described in Examples 1-3.

Figure 17 shows genetic structures thiamine biosynthetic genes and possible mechanisms of riboswitch control. The location and mechanism of the *E. coli* and *B. subtilis* riboswitches are detailed in Examples 2 and 6. The putative TPP riboswitch from *P. secunda* resides immediately upstream from the polyA tail in the cDNA clone of the THIC gene. The putative TPP riboswitch domain in *F. oxysporum* is located in a 5′-UTR intron of the STI35 gene according to the genomic sequence but is absent in the cDNA clone.

Figure 18 shows the L box - a highly conserved sequence and structural domain is present in the 5′-UTRs of Gram-positive and Gram-negative bacterial mRNAs that are related to lysine metabolism. Conserved portions of the L box sequence and secondary structure were identified by alignment of ~~28~~ 31 representative mRNAs as noted (SEQ ID NOS: 29-59). Base pairing potential representing P1 through P5 are ~~individually colored~~ enumerated and set off by boxes. Nucleotides ~~in red~~ shown as lower case letters are conserved in greater than 80% of the examples. The asterisk identifies the representative (*B. subtilis lysC* 5′-UTR) that was examined in this study. Gene names are as annotated in GenBank or were derived by protein sequence similarity. Organism abbreviations are as follows: *Bacillus anthracis* (BA), *Bacillus halodurans* (BH), *Bacillus subtilis* (BS), *Clostridium acetobutylicum* (CA), *Clostridium perfringens* (CP), *Escherichia coli* (EC), *Haemophilus influenzae* (HI), *Oceanobacillus iheyensis* (OI), *Pasteurella multocida* (PM), *Staphylococcus aureus* (SA), *Staphylococcus epidermidis* (SE), *Shigella flexneri* (SF), *Shewanella oneidensis* (SO), *Thermatoga maritima* (TM), Thermoanaerobacter tengcongensis (TT), *Vibrio cholerae* (VC), *Vibrio vulnificus* (VV), *Thermoanaerobacter tengcongensis* (TE).

Figures 19A (SEQ ID NO: 60 and SEQ ID NOS: 400-408), 19B and 19C (SEQ ID NO: 61) show the consensus L box motif from the *lysC* 5′-UTR of *B. subtilis* undergoes allosteric rearrangement in the presence of L-lysine. (*A*) Consensus sequence and structure of the L box domain as derived using a phylogeny of 31 representative sequences from prokaryotic and archaeal organisms (Fig. 18) BA 0845, BA lysA, BA lysP, BH dapA, BH lysC, BH nhaC, BS lysC, BX lysC, CA lysA, CP lysA, CP lysP, EC lysC, HI nhaC, OI dapA, OI nhaC, PM nhaC,

SA lysC, SA lysP, SE lysC, SE lysP, SF lysC, SO lysC, SO nhaC, TM asd, TT lysA, TT pspF, VC lysC, VC nhaC, VC nhaC, VV lysC, VV nhaC, which are represented by SEQ ID NO:29-59, respectively. Nucleotides depicted ~~in red~~ a lower case a, c, t, or g, are present in at least 80% of the representatives, open circles identify nucleotide positions of variable identity, and ~~tan~~ dashed lines denote variable nucleotide identity and chain length. Figure 19B shows sequence, secondary structure model, and lysine-induced structural modulation of the *lysC* 5′-UTR of *B. subtilis*. An additional 94 nucleotides (not depicted) reside between nucleotide 237 and the AUG start codon. Structural modulation sites (~~red-encircled nucleotides~~ nucleotides enclosed in squares) were established using 237 *lysC* RNA by monitoring spontaneous RNA cleavage as depicted in *C*. Figure 19C shows in-line probing of the 237 *lysC* RNA reveals lysine-induced modulation of RNA structure. Patterns of spontaneous cleavage, revealed by product separation using denaturing 10% polyacrylamide gel electrophoresis (PAGE), are altered at four major sites (denoted 1 through 4) in the presence (+) of 10 μM L-lysine (L) relative to that observed in the absence (−) of lysine. T1, $^{-}$OH and NR represent partial digest with RNase T1, partial digest with alkali, and no reaction, respectively. Selected bands in the T1 lane (G-specific cleavage) are identified by nucleotide position. See Methods for experimental details.

Figures 20A, 20B, 20C, 20D and 20E show the molecular recognition characteristics of the lysine aptamer and the use of caged lysine. Figure 20A shows the chemical structures of L-lysine, D-lysine and nine closely-related analogs. Small circles represent chiral carbon centers wherein the enantiomeric configuration is defined for each compound. ~~Shaded~~ Encircled atoms identify chemical differences between L-lysine and the analog depicted. Figure 20B shows in-line probing analysis of the 179 *lysC* RNA in the absence (−) of ligand, or in the presence of 10 μM L-lysine or 100 μM of various analogs as indicated for each lane. For each lane, the relative extent of spontaneous cleavage at site 3 is compared to that of the zone of constant cleavage immediately below this site, where a cleavage ratio significantly below ~1.5 reflects modulation. Figure 20C shows a schematic representation of dipeptide digestion by hydrochloric acid. All dipeptide~~dipetide~~ forms are expected to be incapable of binding the lysine aptamer (*inactive*), while lysine-containing dipeptides should induce conformational changes in the aptamer (*active*) upon acid digestion. Figure 20D shows in-line probing analysis of the 179 *lysC* RNA in the

absence of lysine (–) or in the presence of various amino acids and dipeptides. Underlined lanes carry dipeptide preparations that were pretreated with HCl as depicted in a. Figure 20E shows the fraction of spontaneous cleavage at site 3 in d is plotted after normalization to the extent of processing in the absence of added ligand.

Figures 21A, 21B, 21C and 21D show determination of the dissociation constant and stoichiometry for L-lysine binding to the 179 *lysC* RNA. Figure 21A shows in-line probing with increasing concentrations of L-lysine ranging from 3 nM to 3 mM. Details are as defined for Fig. 19C. Figure 20B shows a plot depicting the normalized fraction of RNA undergoing spontaneous cleavage versus the concentration of amino acid for sites 1 through 3. The dashed line identifies the concentration of L-lysine required to bring about half-maximal structural modulation, which indicates the apparent K_D for ligand binding. Figure 20C shows the 179 *lysC* RNA (10 μM) shifts the equilibrium of tritiated L-lysine (50 nM) in an equilibrium dialysis chamber. To investigate competitive binding, unlabeled L- (L) and D-lysine (D), or L-ornithine (5) were added to a final concentration of 50 μM each to one chamber of a pre-equilibrated assay as indicated. Figure 21D shows a scatchard analysis of L-lysine binding by the 179 *lysC* RNA. The variable r represents the ratio of bound ligand concentration versus the total RNA concentration and the variable $[L_F]$ represents the concentration of free ligand.

Figures 22A, 22B and 22C show the *B. subtilis lysC* riboswitch and its mechanism for metabolite-induced transcription termination. Figure 22A shows a sequence and repressed-state model for the *lysC* riboswitch secondary structure (SEQ ID NO: 62). The encircled nucleotides ~~highlighted in orange~~ identify the putative anti-terminator interaction that could form in the absence of L-lysine. Boxed nucleotides identify sites of disruption (M1) and compensatory mutations for the terminator stem (M2) and for the terminator and anti-terminator stems (M3). Nucleotides ~~in shaded in light blue~~ enclosed in squares identify some of the positions where mutations exhibit *lysC* derepression that were reported previously (Vold et al. 1975; Lu et al. 1992). Figure 22B shows *In vitro* transcription assays conducted in the absence (–) or presence (+) of 10 mM L-lysine or other analogs as indicated. FL and T identify the full-length and terminated transcripts, respectively. The percent of the terminated RNAs relative to the total terminated and full-length transcripts are provided for each lane (% term.). Figure 22C shows *In*

vivo expression of a β-galactosidase reporter gene fused to wild-type (WT), G39A and G40A mutant *lysC* 5′-UTR fragments. Media conditions are as follows: I, normal medium (0.27 mM lysine); II, minimal medium (0.012 mM); III, lysine-supplemented minimal medium (1 mM); IV, lysine hydroxamate-supplemented (medium II plus 1 mM lysine hydroxamate) minimal media; V, thiosine-supplemented (medium II plus 1 mM thiosine) minimal medium.

Figure 23 shows that a highly conserved domain is present in the 5′-UTR of certain gram-positive and gram-negative bacterial mRNAs. Depicted is an alignment of 32 representative mRNA domains from bacteria that conform to the G box consensus sequence BH1-guaA, BH2-[pbuG], BH3-purE, BH4-ssnA, BH5-[xpt], BS1-[pbuG], BS2-purE, BS3-xpt, BS4-yxjA, BS5-ydhL, CA1-uraA, CA2-[pbuG], CA3-guaB, CP1-xpt, CP2-uapC, CP3-guaB, CP4-add, FN1-purQ, LL1-xpt, LM1-[pbuG], LM2-[xpt], OI1-guaA, OI2-[pbuG], OI3-purE, OI4-[xpt], SA1-xpr, TSE1-[xpt], STA1-xpt, STPY1-xpt, STPN-xpt, TE1-[pbuG], VV1-add, which are represented by SEQ ID NO: 63-94 respectively. ~~Regions shaded orange, blue and purple~~ Enclosed and enumerated regions identify base-pairing potential of stems P1, P2, and P3, respectively. Nucleotides ~~in red~~ shown as lower case letters are conserved in greater than 90% of the examples. The asterisk identifies the representative (*xpt-pbuX* 5′-UTR) that was examined in this study. It is important to note that three representatives (BS5, CP4 and VV1) that carry a C to U mutation in the conserved core (in the P3-P1 junction) appear to be adenine-specific riboswitches (unpublished observations). Gene names are as annotated in GenBank, the SubtiList database, or based on protein similarity searches (brackets). Organisms abbreviations are as follows: *Bacillus halodurans* (BH), *Bacillus subtilis* (BS), *Clostridium acetobutylicum* (CA), *Clostridium perfringens* (CP), *Fusobacterium nucleatum* (FN), *Lactococcus lactis* (LL), *Listeria monocytogenes* (LM), *Oceanobacillus iheyensis* (OI), *Staphylococcus aureus* (SA), *Staphylococcus epidermidis* (SE), *Streptococcus agalactiae* (STA), *Streptococcus pyogenes* (STPY), *Streptococcus pneumoniae* (STPN), *Thermoanaerobacter tengcongensis* (TE), and *Vibrio vulnificus* (VV).

Figures 24A, 24B and 24C show the G box RNA of the *xpt-pbuX* mRNA in *B. subtilis* responds allosterically to guanine. Figure 24A shows the consensus sequence and secondary model for the G box RNA domain that resides in the 5′ UTR of genes that are largely involved in

purine metabolism (SEQ ID NO: 95). Phylogenetic analysis is consistent with the formation of a three-stem (P1 through P3) junction. Nucleotides depicted ~~in red~~ shown as lower case letters and ~~black~~ capitals are present in greater than 90% and 80% of the representatives examined, respectively (Figure 23). Encircled nucleotides exhibit base complementation, which might indicate the formation of a pseudoknot. Figure 24B shows sequence and ligand-induced structural alterations of the 5′-UTR of the *xpt-pbuX* transcriptional unit (SEQ ID NO: 96). The putative anti-terminator interaction is ~~highlighted in orange~~ represented by the boxes. Nucleotides that undergo structural alteration as determined by in-line probing (from C) are identified with ~~red circles~~ squares. The 93 *xpt* fragment (boxed) of the 201 *xpt* RNA retains guanine-binding function. Asterisks denote alterations to the RNA sequence that facilitate *in vitro* transcription (5′ terminus) or that generate a restriction site (3′ terminus). Nucleotide numbers begin at the first nucleotide of the natural transcription start site. The translation start codon begins at position 186. Figure 24C shows guanine and related purines selectively induce structural modulation of the 93 *xpt* mRNA fragment. Precursor RNAs (Pre; 5′ ^{32}P-labeled) were subjected to in-line probing by incubation for 40 hr in the absence (–) or presence of guanine, hypoxanthine, xanthine and adenine as indicated by G, H, X and A, respectively. Lanes designated NR, T1 and $^{-}$OH contain RNA that was not reacted, subjected to partial digestion with RNase T1 (G-specific cleavage), or subjected to partial alkaline digestion, respectively. Selected bands corresponding to G-specific cleavage are identified. Regions 1 through 4 identify major sites of ligand-induced modulation of spontaneous RNA cleavage.

Figures 25A and 25B show the 201 xpt mRNA Leader Binds Guanine with High Affinity. Figure 25A shows in-line probing reveals that spontaneous RNA cleavage of the 201 *xpt* RNA at four regions decreases with increasing guanine concentrations. Only those locations of the PAGE image corresponding to the four regions of modulation as indicated in Figure 25C are depicted. Other details and notations are as described in the legend to Figure 25C. Figure 25B shows a plot depicting the normalized fraction of RNA that experienced spontaneous cleavage versus the concentration of guanine for modulated regions 1 through 4 in Figure 25A. Fraction cleaved values were normalized to the maximum cleavage measured in the absence of

guanine and to the minimum cleavage measured in the presence of 10 μM guanine. The apparent K_D value (less than or equal to 5 nM) reflects the limits of detection for these assay conditions.

Figures 26A, 26B and 26C show a molecular discrimination by the guanine-binding aptamer of the *xpt-pbuX* mRNA. Figure 26A shows the chemical structures and apparent K_D values for guanine, hypoxanthine and xanthine (active natural regulators of *xpt-pbuX* genetic expression in *B. subtilis*) versus that of adenine (inactive). Differences in chemical structure relative to guanine are ~~shaded pink~~ encircled. K_D values were established as shown in Figure 26 with the 201 *xpt* RNA. Numbers on guanine represent the positions of the ring nitrogen atoms. Figure 26B shows chemical structures and K_D values for various analogs of guanine reveal that all alterations of this purine cause a loss of binding affinity. Open circles identify K_D values that most likely are significantly higher than indicated, as concentrations of analog above 500 μM were not examined in this analysis. The apparent K_D values of G, H, X and A as indicated are plotted as triangles ~~red triangles~~ for comparison. Figure 26C shows a schematic representation of the molecular recognition features of the guanine aptamer in 201 *xpt*. Hydrogen bond formation at position 9 of guanine is expected because guanosine (K_D > 100 μM) and inosine (K_D > 100 μM), which are 9-ribosyl derivatives of guanine and hypoxanthine, respectively, do not exhibit measurable binding (see Figure 27).

Figures 27A and 27B show confirmation of guanine binding specificity by equilibrium dialysis. Figure 27A shows an equilibrium dialysis strategy was used to confirm that *in vitro*-transcribed *93 xpt* RNAs bind to guanine and can discriminate against various analogs. Each data point was generated by adding ^{3}H-guanine to chamber *a*, which is separated from RNA and other analogs by a dialysis membrane with a molecular weight cut-off (MWCO) of 5,000 daltons. Left: If no guanine binding sites are present in chamber *b*, or if an excess of unlabeled competitor is present, then no shift in the distribution of tritium is expected. Right: If an excess of guanine-binding RNAs are present in chamber *b*, and if no competitor is present, then a substantial shift in the distribution of tritium towards chamber *b* is expected. Figure 27B shows the 93 *xpt* RNA can shift the distribution of ^{3}H-guanine in an equilibrium dialysis apparatus, while analogs of guanine are poor competitors. The plot depicts the fraction of counts per minute (cpm) of tritium in chamber *b* relative to the total amount of cpm counted from both chambers. A value of ~0.5 is

expected if no shift occurs, as is the case when RNA is absent (none), or in the presence of excess unlabeled competitor (G). A value approaching 1 is expected if the majority of ^{3}H-guanine is bound by the RNA in chamber *b* in the absence (–) of unlabeled analog, or in the presence of unlabeled analogs that do not serve as effective competitors under the assay conditions (100 nM ^{3}H-guanine, 300 nM RNA, 500 nM analog). Ino and Gua represents inosine and guanosine, respectively.

Figures 28A, 28B, 28C and 28D show the binding and genetic control functions of variant guanine riboswitches. Figure 28A shows mutations used to examine the importance of various structural features of the guanine aptamer domain (SEQ ID NO: 97). Figure 28B shows examination of the binding function of aptamer variants by equilibrium dialysis. WT designates the wild-type 93 *xpt* construct. Details are as described for Figure 27. Figure 28C shows genetic modulation of a β-galactosidase reporter gene upon the introduction of various purines as indicated. Figure 28D shows regulation of β-galactosidase reporter gene expression by WT and mutants M1 through M7. Open and filled bars represent enzyme activity generated when growing cells in the absence and presence of guanine, respectively.

Figures 29A and 29B show that riboswitches participate in fundamental genetic control. Figure 29A schematic representations of the seven known riboswitches and the metabolites they sense. The secondary structure models were obtained as follows: coenzyme B_{12} (see Example 1); TPP (see Example 2); FMN (see Example 3), SAM (see Example 7); guanine (see Example 6); lysine (see Example 5); adenine (see Example 8). Coenzyme B_{12} is depicted in exploded form wherein *a*, *b* and *c* designate covalent attachment sites between fragments. Figure 29B shows a genetic map of *B. subtilis* riboswitch regulons and their positions on the bacterial chromosome. Genes are controlled by riboswitches as identified by ~~matching color~~ matching numbers. All nomenclature is derived from the SubtiList database release R16.1 (Moszer, I., et al., 1995, Microbiol. *141*, 261-268) except for *metI* and *metC*, which are recent designations (Auger, S., et al., 2002, Microbiol. *148*, 507-518).

Figures 30A, 30B and 30C show the S Box is a structured RNA domain that binds SAM. (A) Consensus sequence and secondary-structure model of the S box domain derived from 107 bacterial representatives (SEQ ID NO: 98 and SEQ ID NOS: 409-410). Lower case letter and

~~Red and black~~ capital letter positions identify nucleotides whose identity as depicted is conserved in greater than 90% or 80% of the representative S box RNAs, respectively. R, Y, and N represent purine, pyrimidine, and any nucleotide, respectively. P1 through P4 identify conserved base pairing. ~~Encircled~~ Enclosed nucleotides identify a putative pseudoknot interaction. Figure 30B shows a sequence and secondary structure model for the 251 *yitJ* mRNA fragment (SEQ ID NO: 99). Sites of structural modulation upon introduction of SAM are depicted as described. Nucleotide 1 corresponds to the putative transcriptional start site. Asterisks identify nucleotides that were added to the construct to permit efficient transcription *in vitro*. The first nucleotide of the AUG start codon is 212 (not shown). Other notations are as described in *a*. Figure 30C shows the spontaneous cleavage patterns of 251 *yitJ* (~1 nM 5′ ^{32}P-labeled) RNA incubated for ~40 hr at 25°C in 50 mM Tris-HCl (pH 8.3 at 25°C), 20 mM $MgCl_2$, 100 mM KCl, and without (–) or with methionine or SAM as indicated for each lane. NR, T1 and $^{-}$OH represent no reaction, partial digest with RNase T1, and partial digest with alkali, respectively. Certain fragment bands corresponding to T1 digestion (cleaves after G residues) are depicted. ~~Blue arrowheads~~ Arrowheads identify positions of significant modulation of spontaneous cleavage, and the numbered sites were used for quantitation (see Fig. 31*b*). Experimental procedures are similar to those described Examples 1-3.

Figures 31A, 31B and 31C show the binding affinity and molecular discrimination by a SAM-binding RNA. Figure 31A shows the chemical structures of various compounds used to probe the binding characteristics of the SAM *yitJ* riboswitch. Other than methionine, each compound as depicted is coupled to an adenosyl moiety ([A]; inset) coupled via the 5′ carbon (as signified by R). Figure 31B Left: The K_D of 251 *yitJ* for SAM was determined by plotting the normalized fraction of RNA cleaved at regions 1 through 6 (see Fig. 30*c*) versus the logarithm of the concentration of SAM in molar units. The dashed line indicates the concentration needed to induce half maximal modulation of cleavage activity. Right: K_D values for SAM and various analogs as determined by this method. Figure 31C shows molecular discrimination determined by equilibrium dialysis. Assays employed 100 nM of *S*-adenosyl-L-methionine-methyl-^{3}H (^{3}H-SAM; 14.5 μCi mmol^{-1}; ~7,000 cpm) added to side *A* of an equilibrium dialysis chamber (*1, 2*), and were conducted in the absence (none) or the presence of 3 μM RNA on the *B* side of the

chamber as indicated. Equilibrations were carried out for ~10 hr in the absence (–) of unlabeled analogs, and then were subsequently incubated in the presence of 25 μM unlabeled compounds (added to side *B*) as indicated. M1 is a variant of 124 *yitJ* that carries disruptive mutations in the junction between stems P1 and P2 (Fig. 32*a*). Line at a cpm ratio of 1 identifies the bar height expected if a shift in ^{3}H-SAM has not occurred. Additional experimental details are similar to those described in Examples 1 and 2.

Figures 32A, 32B and 32C show the effects of RNA mutations on SAM binding and genetic control. Figure 32A shows the sequence and secondary structure model for the 124 *yitJ* RNA (SEQ ID NO: 100). Mutations M1 through M9 were generated in plasmids containing fusions of the *yitJ* 5′-UTR upstream from a *lacZ* reporter gene. Templates for preparation of mutant RNAs for *in vitro* studies were then created by PCR, and the mutant DNA constructs were integrated into the chromosome for *in vivo* studies. See Methods for experimental details. Figure 32B shows the analysis of SAM-binding function by equilibrium dialysis in the presence of wild-type (WT) and mutant RNAs as denoted. Details are described in the legend to Fig. 31*c*, except that 300 nM RNA was used and all assays were conducted without the addition of unlabeled analogs. Figure 32C shows *In vivo* control of β-galactosidase expression in *B. subtilis* cells transformed with various riboswitch constructs as indicated. β-galactosidase activities were measured as described in Example 2. Cells were grown in glucose minimal media in 0.75 μg mL^{-1} methionine (–) 50 μg mL^{-1} methionine (+). M6 through M9 were not examined *in vivo*.

Figures 33A and 33B show metabolite-induced transcription termination of several mRNAs that carry a SAM riboswitch. Figure 33A shows *In vitro* transcription using T7 RNA polymerase results in increased termination of four mRNA leader sequences. Reactions were conducted in the absence (–) or presence (+) of 50 μM of the effector as indicated for each lane. For example, the *metI* template includes the 5′ UTR and coding sequences through mRNA position 242, while the termination site is expected to occur at position 189. Below each gel is indicated the percentage of transcription termination (T) at the expected location relative the total amount of expected termination plus full length RNA (FL). Figure 33B shows sequence and structural model for the *metI* riboswitch in two structural states (SEQ ID NO: 101). ~~Green and pink residues~~ Residues shown in hexagons and squares correspond to the P1 (anti-anti-

terminator) and the terminator stems, respectively. The ~~orange~~ encircled residues correspond to the anti-terminator stem. Sequences boxed in black define ~~red define~~ the location and identity of mutations used to examine the proposed mechanism of genetic control. Gel: Analysis of mutant *metI* riboswitches wherein disruptive (M*a*, M*ab* and M*c*) or the corresponding compensatory mutations (M*abc*) have been inserted. The *metI* mutant templates and wild-type control template (WT) are identical to the templates used in A, except that the FL product is 220 nucleotides. Other notations are as describe in A.

Figures 34A and 34B show Bacilli species *subtilis* and *anthrasis* bind SAM with different affinities. Figure 34A shows structural modulation of the *B. subtilis cysH* aptamer as determined by in-line probing (SEQ ID NO: 102). Inset: Apparent K_D values determined by monitoring structural modulation over a range of SAM or SAM analog concentrations. Two G residues (asterisks) were included at the 5' terminus of the RNA construct to facilitate *in vitro* transcription. Nucleotide numbers are given relative to the putative transcription start site. In-line probing was conducted with an RNA extending to nucleotide 117, while the remainder of the RNA is shown to depict the putative transcription terminator stem. Experiments were similar to those described in Fig. 30*b* and Fig. 31*b*. See the legend for Fig. 30*b* for details. Figure 34B shows structural modulation of the *B. subtilis cysH* aptamer as determined by in-line probing (SEQ ID NO: 103). The transcription start point of the *B. anthracis cysH* mRNA has not been determined, and so numbering of nucleotides begins immediately after the two inserted G residues (asterisks). In-line probing was conducted with an RNA extending to nucleotide 112. See *a* for additional details.

Figures 35A, 35B and 35C show guanine- and adenine-specific riboswitches. Figure 35a shows sequence and structural features of the two guanine-specific (*purE* and *xpt*) and three adenine-specific aptamer domains that are examined in this study BS2-purE, BS3-xpt, BS5-ydhL, CP4-add, VV1-add, which are represented by SEQ ID NOS: 104-108, respectively. P1 through P3 identify the three base-paired stems comprising the secondary structure of the aptamer domain. Lowercase nucleotides ~~Red nucleotides~~ identify positions whose base identity is conserved in greater than 90% of representatives in the phylogeny[1]. The arrow identifies a nucleotide within the conserved core of the aptamer that is a determinant of ligand specificity.

BS, CP and VV designate *B. subtilis*, *Clostridium perfringens* and *Vibrio vulnificus*, respectively. Figure 35b shows sequence and secondary structure of the *xpt* and *ydhL* aptamers (SEQ ID NO: 109). ~~Green~~ Encircled nucleotides identify positions within the *ydhL* aptamer that differ from those in the *xpt* aptamer. The sequence disclosed in Figure 35c is SEQ ID NO: 110. Nucleotides in xpt are numbered as described in Example 6. Other notations are as described in A.

Figures 36A, 36B, 36C, 36D and 36E show the ligand specificity of five G box RNAs. (a through e) In-line probing assays for the conserved aptamer domains as labeled. NR, T1 and $^{-}$OH identify marker lanes wherein precursor RNAs (Pre) were not incubated, or were partially digested with RNase T1 or alkali, respectively. Selected bands corresponding to RNase T1 digestion (cleavage 3′ relative to guanidyl residues) are labeled for each RNA. RNAs were incubated for 40 hr in the absence of ligand (–), or in the presence of 1 μM guanine (G) or adenine (A). Large arrowheads identify sites of substantial change in cleavage pattern that is due to the addition of a particular ligand. See Methods for additional details.

Figures 37A and 37 B show the binding affinity of the *ydhL* aptamer for adenine. Figure 37a shows the in-line probing assay for the 80 *ydhL* RNA at various concentrations of adenine. For each lane, sites 1 through 4 were quantitated and the fraction of RNA cleaved was used to determine the apparent K_D. Figure 37b shows a plot of the normalized fraction of RNA that has undergone spontaneous cleavage at sites 1 through 4 versus the concentration of adenine. See Example 8 for additional details.

Figures 38A and 38B show the specificity of molecular recognition by the adenine aptamer from *ydhL*. Figure 38a Top: Chemical structures of adenine, guanine and other purine analogs that exhibit measurable binding to the 80 *ydhL* RNA. Chemical changes relative to 2,6-DAP, which is the tightest-binding compound, are ~~highlighted in pink~~ encircled. Bottom left: Plot of the apparent K_D values for various purines. Bottom right: Model for the chemical features on adenine that serve as molecular recognition contacts for *ydhL*. Note that the importance of N7 and N9 has not been determined. Encircled arrow indicated that a contact could exist if a hydrogen bond donor is appended to C2. Figure 38b shows chemical structures of various purines that are not bound by the 80 *ydhL* RNA (K_D values poorer than 300 μM).

Figures 39A, 39B, 39C and 39D show interconversion of guanine- and adenine-specific aptamers. Figure 39a Left: Plot of the normalized fraction of wild-type 93 *xpt* RNA cleavage product for a given site versus the logarithm of the concentration of ligand present during incubation in an in-line probing assay. Cleavage products monitored for modulation correspond to site 3 (Fig. 37a). Right: Plot of the fraction of the total counts per minute (cpm) present in chamber B relative to the total counts per minute from sides *A* and *B* of an equilibrium dialysis chamber. Value of ~0.5 indicate an equal distribution of ligand (no binding) while values of ~1 indicate that most of the ligand is bound to the RNA within side *B* of the chamber. (b, c, d) In-line probing plots and equilibrium dialysis plots for 93 *xpt* (C to U mutation), 80 *ydhL*, and 80 *ydhL* (U to C mutation), respectively. Details are describe in a, or are described in the Example 8.

Figures 40A, 40B, 40C, 40D and 40E show a model for the genetic control of *ydhL* by an adenine riboswitch and its function as a gene-activating element. Figure 40a sequence of the adenine riboswitch from *B. subtilis ydhL* and secondary structure models for the 'ON' and 'OFF' states for gene regulation (SEQ ID NO: 111). Figure 40b *In vivo* function of the wild-type *ydhL* riboswitch and of a variant form as determined by fusion to a β-galactosidase reporter gene.

Figures 41A-41JB show the sequence and types of riboswitches Bs01, Bs02, Bs03, Bs04, Bs05, Bs06, Bs07, Bs08, Bs09, Bs10, Bs11, Bh01, Bh02, Bh03, Bh04, Bh05, Oi01, Oi02, Oi03, Oi04, Oi05, Oi06, Oi07, Oi10, Oi08, Oi09, Oi10, Oi11, Oi12, Oi13, Ca01, Ca02, Ca03, Ca04, Ca05, Ca06, Ca07, Cp01, Cp02, Lm01, Lm02, Lm03, Lm04, Lm05, Lm06, Lm07, Li01, Li02, Li03, Li04, Li05, Li06, Li07, Sa01, Sa02, Sa03, Sa04, Sc01, Ct01, Tt01, Tt02, Tt03, Fn01, Fn02, Dr01, Dr02, Xa01, Xc01, Se01, Se02, Gs01, Gs02, Ba01, Ba02, Ba03, Ba04, Ba05, Ba06, Ba07, Ba08, Ba09, Ba10, Ba11, Ba12, Ba13, Ba14, Ba15, Ba16, Ba17, Bc01, Bc02, Bc03, Bc04, Bc05, Bc06, Bc07, Bc08, Bc09, Bc10, Bc11, Bc12, Bc13, Bc14, Bc15, Bc16, Bc17, Bc18, Atu01, Atu02, Atu03, Atu04, Atu05, Atu06, Bha01, Bha02, Bha03, Bha04, Bsu01, Bja01, Bja02, Bja03, Bja04, Bja05, Bme01, Bme02, Bme03, Bme04, Ccr01, Ccr02, Cte01, Cte02, Cte03, Cte04, Cte05, Cac01, Cac02, Cpe01, Cpe02, Cpe03, Cpe04, Eco01, Fnu01, Lig01, Lmo01, Mlo01, Mlo02, Mlo03, Mlo04, Mlo05, Mlo06, Mle01, Mtu01, Mtu02, Pae01, Pae02, Pae03, Pae04, Ppu01, Ppu02, Ppu03, Ppu04, Rso01, Sme01, Sme02, Sme03, Sme04, Sme05, Sco01, Sco02,

Sco03, Sco04, Sco05, Sfl01, Son01, Son02, Sti01, Sti02, Tma01, Tte01, Tte02, Vch01, Vvu01, Xac01, Xax01, Ype01, Aca01, Avi01, Bfr01, Bmg01, Lma01, Pfr01, Rca01, Rca02, Rca03, Rsp01, Sbi01, Sgi01, Svi01, Zmo01, Zmo02, NC_002570.1/648448-648540, NC_002570.1/650317-650406, NC_002570.1/676483-676572, NC_002570.1/806882-806965, NC_002570.1/1593067-1592976, NC_000964.1/693955-694038, NC_000964.1/697886-697976, NC_000964.1/2319120-2319031, NC_000964.1/4004319-4004410, NC_003030.1/1002184-1002270, NC_003030.1/2904259-2904168, NC_003030.1/2824539-2824454, NC_003366.1/422828-422924, NC_003366.1/512410-512323, NC_003366.1/2617892-2617807, NC_003454.1/1645257-1645173, NC_002662.1/1159519-1159604, NC_003210.1/610773-610679, NC_003210.1/1958601-1958511, NC_004193.1/760480-760571, NC_004193.1/769695-769781, NC_004193.1/786775-786863, NC_004193.1/1103947-1104044, NC_002745.1/430771-430861, NC_004461.1/2432384-2432294, NC_004116.1/1093950-1093860, NC_002737.1/930757-930842, NC_003028.1/1754791-1754878, NC_003869.1/586372-586463, NC_000964.1/626134-626051, NC_003366.1/2870819-2870732, NC_004460.1/504378-504467, Bha_LysC, Bha_dapA, Bha_nhaC, Bsu_LysC, Cac_lysA, Cpe_nhaC, Cpe_lysA, Cpe_lysP, Eco_lysC, Hin_nhaC, Oih_dapA, Oih_nhaC, Pmu_nhaC, Sau_lysC, Sau_lysP, Sep_lysC, Sep_lysP, Sfl_lysC, Son_lysC, Son_nhaC, Tma_asd, Tte_lysA, Tte_pspF, Vch_lysC, Vch_nhaC, Vch_nhaC, 2Vvu_lysC, Vvu_nhaC, Cons, Cons and Consensus, which are represented by SEQ ID NO:112-374, respectively.

Please replace the paragraph bridging pages 85 and 86 with the following paragraph:

Metabolite-dependent conformational changes in the 202-nucleotide leader sequence of the *btuB* mRNA. Figure 1A: Separation of spontaneous RNA-cleavage products of the *btuB* leader using denaturing 10% polyacrylamide gel electrophoresis (PAGE). 5'-32p-labeled mRNA leader molecules (arrow) were incubated for 41 hr at 25°C in 20 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.3 at 25°C) in the presence (+) or absence (-) of 20 μM of AdoCbl. Lanes containing RNAs that have undergone no reaction, partial digest with alkali, and partial digest with RNase T1 (G-specific cleavage) are identified by NR, ^{-}OH, and T1, respectively. The location of product bands

corresponding to cleavage after selected guanosine residues are identified by filled arrowheads. ~~Light blue arrowheads~~ Arrowheads labeled 1 through 8 identify eight of the nine locations that exhibit effector-induced structure modulation, which experience an increase or decrease in the rate of spontaneous RNA cleavage. The image was generated using a phosphorimager (Molecular Dynamics), and cleavage yields were quantitated by using ImageQuant software. Figure 1B: Sequence and secondary-structure model for the 202-nucleotide leader sequence of *btuB* mRNA in the presence of AdoCbl. Putative base-paired elements are designated P1 through P9. Complementary nucleotides in the loops of P4 and P9 that have the potential to form a pseudoknot are juxtaposed. Nine specific sites of structure modulation are identified by ~~light blue~~ arrowheads. The asterisks demark the boundaries of the B_{12} box (nucleotides 141-162). The coding region and the 38 nucleotides that reside immediately 5' of the start codon (nucleotides 241-243) were not included in the 202-nucleotide fragment. The 315-nucleotide fragment includes the 202-nucleotide fragment, the remaining 38 nucleotides of the leader sequence, and the first 75 nucleotides of the coding region.

Please replace the paragraph on page 97, lines 8-23, with the following paragraph:

Figure 6B shows a secondary-structure model of 165 *thiM* as predicted by computer modeling (Zuker et al., Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA Biochemistry and Biotechnology* (eds. Barciszewski J. & Clark, B.F.C.) 11-43 (NATO ASI Series, Kluwer Academic Publishers, 1999); Mathews et al., Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288,** 911-940 (1999)) and by the structure probing data depicted in Figure 6A. Spontaneous cleavage characteristics are as noted in the inset. Unmarked nucleotides exhibit a constant but low level of degradation. The truncated 91 *thiM* RNA is boxed and the *thi* box element (Miranda-Rios et al., A conserved RNA structure (*thi* box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc. Natl. Acad. Sci. USA* **98,** 9736-9741 (2001)) is shaded ~~light blue~~. Nucleotides ~~highlighted in orange~~ enclosed in boxes identify an alternative pairing, designated P8*. The RNA carries two mutations (G156A

and U157C) relative to wild type that were introduced in a non-essential portion of the construct to form a restriction site for cloning, while all RNAs carry two 5′-terminal G residues to facilitate *in vitro* transcription.

Please replace the paragraph on page 97, lines 24-28, with the following paragraph:

Figure 6C shows TPP-dependent modulation of the spontaneous cleavage of 240 *thiC* RNA. Reactions were conducted and analyzed as described in above for Figure 6A. Figure 6D shows a secondary-structure model of 240 *thiC*. Base-paired elements that are similar to those of *thiM* are labeled P1 through P5. The truncated RNA 111 *thiC* is boxed. Nucleotides ~~highlighted in orange~~ enclosed in boxes identify an alternative pairing.

Please replace the paragraph bridging pages 97 and 98 with the following paragraph:

Figure 7A shows the extent of spontaneous modulation of RNA cleavage at several sites within 165 *thiM* (left) and 240 *thiC* (right) plotted for different concentrations (*c*) of TPP. ~~Red arrows~~ Arrows reflect the estimated concentration of TPP needed to attain half maximal modulation of RNA (apparent K_D). Figure 7B shows the logarithm of the apparent K_D values plotted for both RNAs with TPP, TP and thiamine as indicated. The boxed data was generated using TPP with the truncated RNAs 91 *thiM* and 111 *thiC*. Figure 7C shows that patterns of spontaneous cleavage of 165 *thiM* differ between thiamine and TPP ligands as depicted by PAGE analysis (left) and as reflected by graphs (right) representing the relative phosphorimager counts for the three lanes as indicated. Details for the RNA probing analysis are similar to those described above in connection with Figure 6A. The graphs were generated by ImageQuant software.

Please replace the paragraph bridging pages 98 and 99 with the following paragraph:

Figure 9A shows mutations present in constructs M1 through M8 relative to the 165 *thiM* RNA. P8* is a putative base-paired element between portions (~~orange~~shaded) of the P1 and P8 stems. Figure 9B (top) shows *in vitro* ligand-binding and genetic control functions of the wild-type (WT), M1 and M2 RNAs as reflected by PAGE analysis of in-line probing experiments (10 μM TPP) and by β-galactosidase expression assays. Labels on PAGE gels are as described above in connection with Figure 6A. Bars represent the levels of gene expression in the presence (+) and the absence (-) of TPP in the culture medium. Figure 9C is a summary of similar analyses of WT through M9 is presented in table form. The SD status "n.d." (not determined) indicates either that the level of spontaneous cleavage detected in the absence and presence of TPP is near the limit of detection (M6, M7 and M8) or that the region adopts an atypical structure (M9) compared to WT.

Please replace the paragraph on page 122, lines 7-15, with the following paragraph:

L box domains were identified by sequence similarity to the *B. subtilis lysC* 5′-UTR. Ultimately, the program was used to search for degenerate matches to the pattern (WAGAGGNGC [10] A [3] RKTA [50] RRGR [10] CCGARR [40] GG [13] VAA [13] YTGTCA [36] TGRWG [2] CTWY) (SEQ ID NO: 376), however, less complete versions of this pattern were used with iterative refinements to identify the consensus sequence and structure of the L box motif. Bracketed numbers are variable gaps with constrained maximum lengths denoted. Nucleotide notations are as follows: Y = pyrimidine; R = purine; W = A or T; K = G or T; V = A, G or C. Up to six violations of this pattern were permitted when forming the phylogeny depicted in Figure 18.

Please replace the paragraph on page 136, lines 2-15, with the following paragraph:

G box domains were identified by sequence similarity to the *xpt-pbuX* 5′-UTR by conducting a BLASTN search of Genbank using default parameters. These hits were expanded by searching for degenerate matches to the pattern (<<<< [2] TA [6] <<< [2] ATNNGG [2] >>> [5] GTNTCTAC [3] <<<<< [3] CCNNNAA [3] >>>>> [5] >>>>) (SEQ ID NO: 377). Angled

brackets indicate base pairing. Bracketed numbers are variable gaps with constrained maximum lengths denoted. A total of four violations of this pattern were permitted when forming the phylogeny depicted in Figure 23. It is important in this instance to note that only the BS3-xpt domain (that of the *xpt-pbuX* leader) has been shown to bind guanine. It was demonstrated that the molecular specificity of the VV1 representative is for adenine and not guanine (unpublished data). Given the possible trivial means by which a guanine-binding RNA aptamer might be altered to bind adenine (e.g. a C to U change if the C residue is used by the aptamer to make a Watson-Crick-pairing interaction with guanine), it cannot be ruled out that other representatives also have altered molecular recognition.

Please replace the paragraph bridging pages 164 and 165 with the following paragraph:

Additional riboswitches were found based on published alignments and secondary structures (Grundy, F.J. & Henkin, T.M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol. Microbiol.* **30**, 737-749 (1998)) using the SequenceSniffer program. This program finds degenerate matches to RNA patterns defined by linked sequence motifs and base pairing constraints. In the alignments, base pairing regions have the identical underline styles or boxes~~colored backgrounds~~ and are labeled as in the corresponding figures discussed in Examples 1-8, with the addition of a putative pseudoknot marked PS. Predicted terminators (short dashed underline~~yellow~~) and start codons (long dashed underline~~green~~) are marked for some sequences. Positions for each sequence in the indicated Genbank record or unfinished genome contig are for the sequence column marked with a circle (•) – the fifth base in stem P1 that is 5' of the aptamer. Start is the offset from the column marked with an asterisk (*) – the sixth base in stem P1 that is 3' of the aptamer – to the start codon of the first gene in the operon. Genes were identified from COGNITOR (Tatusov, R.L., *et al.* The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**, 22-28 (2001)) and PFAM (Bateman, A., *et al.* The Pfam Protein Families Database. *Nucleic Acids Res.* **30**, 276-280 (2002)) database matches to protein sequences annotated in the Genbank records. The standard names from these databases are used when possible (2011 = COG2011; ???? = no matches).

Previous operon designations for *B. subtilis* are given in parentheses (Grundy, F.J. & Henkin, T.M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol. Microbiol.* **30**, 737-749 (1998)). A subset of sequences with <90% pairwise identity between the bases encompassed by stem P1 was selected for determining the consensus sequence. In the consensus sequence, lowercase and uppercase bases indicate >80% and >95% conservation at a position, respectively. Purine (R) and pyrimidine (Y) bases were assigned when no single base had >80% conservation.

Please replace the original drawings (89 sheets) with the 143 sheets of replacement drawings included as an appendix with this Amendment.

Amendments to the Claims

1. (Original) A regulatable gene expression construct comprising a nucleic acid molecule encoding an RNA comprising a riboswitch operably linked to a coding region, wherein the riboswitch regulates expression of the RNA, wherein the riboswitch and coding region are heterologous.

2. (Original) The construct of claim 1 wherein the riboswitch comprises an aptamer domain and an expression platform domain, wherein the aptamer domain and the expression platform domain are heterologous.

3. (Original) The construct of claim 1 wherein the riboswitch comprises an aptamer domain and an expression platform domain, wherein the aptamer domain comprises a P1 stem, wherein the P1 stem comprises an aptamer strand and a control strand, wherein the expression platform domain comprises a regulated strand, wherein the regulated strand, the control strand, or both have been designed to form a stem structure.

4. (Original) A riboswitch, wherein the riboswitch is a non-natural derivative of a naturally-occurring riboswitch.

5. (Original) The riboswitch of claim 4 wherein the riboswitch comprises an aptamer domain and an expression platform domain, wherein the aptamer domain and the expression platform domain are heterologous.

6. (Original) The riboswitch of claim 4 wherein the riboswitch is derived from a naturally-occuring guanine-responsive riboswitch, adenine-responsive riboswitch, lysine-responsive riboswitch, thiamine pyrophosphate-responsive riboswitch, adenosylcobalamin-responsive riboswitch, flavin mononucleotide-responsive riboswitch, or a S-adenosylmethionine-responsive riboswitch.

7. (Original) The riboswitch of claim 4 wherein the riboswitch is activated by a trigger molecule, wherein the riboswitch produces a signal when activated by the trigger molecule.

8. (Withdrawn) A method of detecting a compound of interest, the method comprising bringing into contact a sample and a riboswitch, wherein the riboswitch is activated by the compound of interest, wherein the riboswitch produces a signal when activated by the

compound of interest, wherein the riboswitch produces a signal when the sample contains the compound of interest.

9. (Withdrawn) The method of claim 8 wherein the riboswitch changes conformation when activated by the compound of interest, wherein the change in conformation produces a signal via a conformation dependent label.

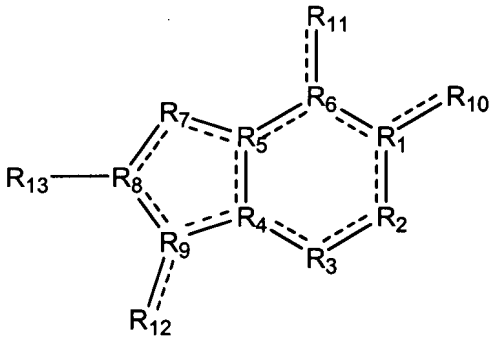
10. (Withdrawn) The method of claim 8 wherein the riboswitch changes conformation when activated by the compound of interest, wherein the change in conformation causes a change in expression of an RNA linked to the riboswitch, wherein the change in expression produces a signal.

11. (Withdrawn) The method of claim 10 wherein the signal is produced by a reporter protein expressed from the RNA linked to the riboswitch.

12. (Withdrawn) A method of inhibiting gene expression, the method comprising

bringing into contact a compound and a cell,

wherein the compound has the structure



wherein, when the compound is bound to a guanine-responsive riboswitch, R_7 serves as a hydrogen bond acceptor, R_{10} serves as a hydrogen bond donor, R_{11} serves as a hydrogen bond acceptor, R_{12} serves as a hydrogen bond donor,

wherein R_{13} is H, H_2 or is not present,

wherein R_1, R_2, R_3, R_4, R_5, R_6, R_8, and R_9 are each independently C, N, O, or S,

wherein ------ each independently represent a single or double bond,

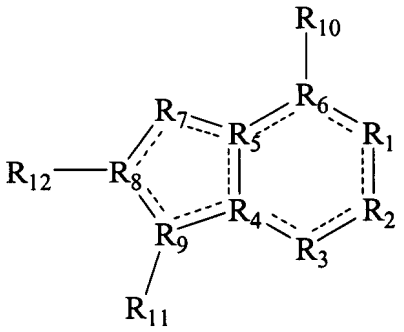
wherein the compound is not guanine, hypoxanthine, or xanthine,

wherein the cell comprises a gene encoding an RNA comprising a guanine-responsive riboswitch, wherein the compound inhibits expression of the gene by binding to the guanine-responsive riboswitch.

13. (Withdrawn) A method of inhibiting gene expression, the method comprising

bringing into contact a compound and a cell,

wherein the compound has the structure



wherein, when the compound is bound to an adenine-responsive riboswitch, R_1, R_3 and R_7 serve as hydrogen bond acceptors, and R_{10} and R_{11} serve as hydrogen bond donors,

wherein R_{12} is H, H_2 or is not present,

wherein R_1, R_2, R_3, R_4, R_5, R_6, R_8, and R_9 are each independently C, N, O, or S,

wherein ------ each independently represent a single or double bond,

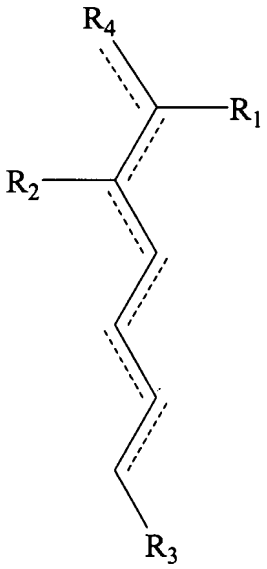
wherein the compound is not adenine, 2,6-diaminopurine, or 2-amino purine,

wherein the cell comprises a gene encoding an RNA comprising an adenine-responsive riboswitch, wherein the compound inhibits expression of the gene by binding to the adenine-responsive riboswitch.

14. (Withdrawn) A method of inhibiting gene expression, the method comprising

bringing into contact a compound and a cell,

wherein the compound has the structure



wherein R_2 and R_3 are each positively charged,

wherein R_1 is negatively charged,

wherein R_4 is C, N, O, or S,

wherein ------ each independently represent a single or double bond,

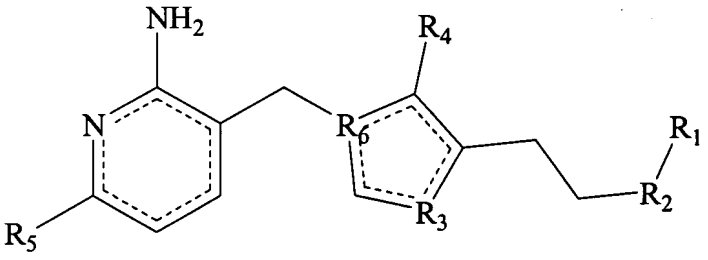
wherein the compound is not lysine,

wherein the cell comprises a gene encoding an RNA comprising a lysine-responsive riboswitch, wherein the compound inhibits expression of the gene by binding to the lysine-responsive riboswitch.

15. (Withdrawn) The method of claim 14 wherein R_2 and R_3 are each NH_3^+ and wherein R_1 is O^-.

16. (Withdrawn) A method of inhibiting gene expression, the method comprising bringing into contact a compound and a cell,

wherein the compound has the structure



wherein R_1 is positively charged,

wherein R_2 and R_3 are each independently C, O, or S,

wherein R_4 is CH_3, NH_2, OH, SH, H or not present,

wherein R_5 is CH_3, NH_2, OH, SH, or H,

wherein R_6 is C or N,

wherein ------ each independently represent a single or double bond,

wherein the compound is not TPP, TP or thiamine,

wherein the cell comprises a gene encoding an RNA comprising a thiamine pyrophosphate -responsive riboswitch, wherein the compound inhibits expression of the gene by binding to the thiamine pyrophosphate-responsive riboswitch.

17. (Withdrawn) The method of claim 16 wherein R_1 is phosphate, diphosphate or triphosphate.

18. (Withdrawn) A method comprising

(a) testing a compound for inhibition of gene expression of a gene encoding an RNA comprising a riboswitch, wherein the inhibition is via the riboswitch,

(b) inhibiting gene expression by bringing into contact a cell and a compound that inhibited gene expression in step (a),

wherein the cell comprises a gene encoding an RNA comprising a riboswitch, wherein the compound inhibits expression of the gene by binding to the riboswitch.

19. (Withdrawn) A method of identifying riboswitches, the method comprising

assess in-line spontaneous cleavage of an RNA molecule in the presence and absence of a compound, wherein the RNA molecule is encoded by a gene regulated by the compound,

wherein a change in the pattern of in-line spontaneous cleavage of the RNA molecule indicates a riboswitch.

20. (New) The construct of claim 2, wherein the expression platform domain comprises an expression regulatory element.

21. (New) The construct of claim 20, wherein the expression regulatory element is selected from the group comprising Shine-Dalgarno sequences, initiation codons, transcription terminators, and stability and processing signals.